ESTIMATION OF OBESITY LEVELS BASED ON EATING HABITS AND PHYSICAL CONDITIONS

Abstract

This study aims to estimate obesity levels by examining the relationships between eating habits, physical activity, and the likelihood of obesity. A regression model is developed using a comprehensive dataset to predict obesity based on key variables.

Dataset

This dataset includes information for predicting obesity levels in individuals from Mexico, Peru, and Colombia, considering their eating habits and physical condition. It consists of 17 attributes and encompasses a total of 2111 records.

The dataset was taken from UCI Machine Learning Repository;

https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition

Methodology

This study utilizes a dataset for estimating obesity levels in individuals from Mexico, Peru, and Colombia based on their eating habits and physical condition. The initial step involves cleaning and preprocessing the data, addressing aspects such as column names, data types, and converting categorical values to numerical ones. Subsequently, data visualization techniques are applied to uncover patterns and relationships within the dataset. Statistical methods, including the use of correlation matrices, are then employed to identify relevant features. To construct a regression model, the dataset is divided into a training set and a test set. Standard scaling is implemented to normalize the data. Following model construction, an evaluation of its performance is conducted to assess its effectiveness in estimating obesity levels.

Before Preprocessing

Gender	Age	Height	Weight	family_histor	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
Female	21	1.62	64	yes	no	2	3	Sometimes	no	2	no	0	1	no	Public_Transportation	Normal_Weight
Female	21	1.52	56	yes	no	3	3	Sometimes	yes	3	yes	3	0	Sometimes	Public_Transportation	Normal_Weight
Male	23	1.8	77	yes	no	2	3	Sometimes	no	2	no	2	1	Frequently	Public_Transportation	Normal_Weight
Male	27	1.8	87	no	no	3	3	Sometimes	no	2	no	2	0	Frequently	Walking	Overweight_Level_I
Male	22	1.78	89.8	no	no	2	1	Sometimes	no	2	no	0	0	Sometimes	Public_Transportation	Overweight_Level_II
Male	29	1.62	53	no	yes	2	3	Sometimes	no	2	no	0	0	Sometimes	Automobile	Normal_Weight
Female	23	1.5	55	yes	yes	3	3	Sometimes	no	2	no	1	0	Sometimes	Motorbike	Normal_Weight
Male	22	1.64	53	no	no	2	3	Sometimes	no	2	no	3	0	Sometimes	Public_Transportation	Normal_Weight
Male	24	1.78	64	yes	yes	3	3	Sometimes	no	2	no	1	1	Frequently	Public_Transportation	Normal_Weight
Male	22	1.72	68	yes	yes	2	3	Sometimes	no	2	no	1	1	no	Public_Transportation	Normal_Weight
Male	26	1.85	105	yes	yes	3	3	Frequently	no	3	no	2	2	Sometimes	Public_Transportation	Obesity_Type_I
Female	21	1.72	80	yes	yes	2	3	Frequently	no	2	yes	2	1	Sometimes	Public_Transportation	Overweight_Level_II

After Preprocessing

1	Age	Height	Weight	mily_History_With_Overwe	High_Caloric_Food_Freque	etable_Consumption_Freque	mber_Of_Main_Meals_D	sumption_Of_Food_Between_	Smoking	Liquid_Intake_Daily	rie_Consumption_Monito	hysical_Activi	e_Using_Technological_Devi	Alcohol_Consumptio	pe_Of_Transportatio	Obesity_Level
2	21	1.62	64	1	0	2	3	1	0	2	0	0	1	0	4	1
3	21	1.52	56	1	0	3	3	1	1	3	1	3	0	1	4	1
4	23	1.8	77	1	0	2	3	1	0	2	0	2	1	2	4	1
5	27	1.8	87	0	0	3	3	1	0	2	0	2	0	2	5	2
6	22	1.78	89.8	0	0	2	1	1	0	2	0	0	0	1	4	2
7	29	1.62	53	0	1	2	3	1	0	2	0	0	0	1	1	1
8	23	1.5	55	1	1	3	3	1	0	2	0	1	0	1	3	1
9	22	1.64	53	0	0	2	3	1	0	2	0	3	0	1	4	1
10	24	1.78	64	1	1	3	3	1	0	2	0	1	1	2	4	1
11	22	1.72	68	1	1	2	3	1	0	2	0	1	1	0	4	1
12	26	1.85	105	1	1	3	3	2	0	3	0	2	2	1	4	3
13	21	1.72	80	1	1	2	3	2	0	2	1	2	1	1	4	2
14	22	1.65	56	0	0	3	3	1	0	3	0	2	0	1	4	1
15	41	1.8	99	0	1	2	3	1	0	2	0	2	1	2	1	3
16	23	1.77	60	1	1	3	1	1	0	1	0	1	1	1	4	1
17	22	1.7	66	1	0	3	3	3	0	2	1	2	1	1	4	1
18	27	1.93	102	1	1	2	1	1	0	1	0	1	0	1	4	2
19	29	1.53	78	0	1	2	1	1	0	2	0	0	0	0	1	3

Model Development

Building a regression model;

Building And Fitting Regression model

In [13]:

```
# Building Multiple Linear Regression model
from sklearn.linear_model import LinearRegression
model = LinearRegression()
```

In [14]:

```
# Fitting Multiple Linear Regression to the Training set
model.fit (X_train_scaled, y_train)
```

Out[14]:

```
LinearRegression()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

Model Evaluation

In [19]:
```
# Calculating the R squared value
from sklearn.metrics import r2_score
r2_score(y_test, y_pred)
```

Out[19]: 0.8701434870082194

In [20]:
```
import seaborn as sns
sns.displot(y_pred-y_test,kind='kde')
```

Out[20]: <seaborn.axisgrid.FacetGrid at 0x152953d30d0>